



ANALYSE DE L'INFORMATION CHIFFRÉE

EXPLOITATION DES DONNÉES DE LA PLATEFORME PARCOURSUP

Prérequis

Utilisation d'un tableur

Références au programme

Situation problème sur l'analyse croisée de couples de caractères

Domaine

Analyse statistique de deux caractères

Tableau croisé d'effectifs

Compétences mathématiques

Chercher : exploitation de données extraites d'un fichier de type *open data*

Représenter : utilisation de graphiques de type diagrammes en barres, diagrammes circulaires

Calculer : calcul d'effectifs, calcul de fréquences marginales, calcul de fréquences conditionnelles

Communiquer : interpréter à l'écrit ou à l'oral les résultats de l'exploitation de données statistiques

Histoire, enjeux, débats

L'*open data* désigne un mouvement, né en Grande-Bretagne et aux États-Unis, d'ouverture et de mise à disposition des données produites et collectées par les services publics (administrations, collectivités locales...).

Retrouvez éducol sur



Les administrations peuvent permettre au public de consulter en ligne une partie de leur base de données : il s'agit de l'*open data* (base de données ouverte). Elles s'engagent alors à respecter le règlement général sur la protection des données (RGPD). La CNIL (Commission nationale de l'informatique et des libertés) et la CADA (Commission d'accès aux documents administratifs) présentent les modalités de mise en ligne et de réutilisation de ces documents : obligations de publication en ligne, contenu des documents publiés, modalités de diffusion en ligne, réutilisation des données diffusées¹... [Source : [site de la CNIL](#)]

Les mathématiques dans le quotidien des élèves

La situation proposée permet de réaliser une analyse des données extraites de la plateforme Parcoursup pour l'année 2021. Les fichiers sources sont de grande taille et contiennent des informations variées comme les vœux de poursuite d'études et de réorientation dans l'enseignement supérieur et les réponses des établissements. Cette variété permet d'éclairer les élèves sur des questions d'ordre sociologique, de développer leur sens critique et de les conduire vers une réflexion sur les parcours d'orientation.

Intentions pédagogiques

L'activité proposée a pour objectif d'illustrer l'analyse de l'information chiffrée dans le contexte de l'orientation post-bac. Pour ce faire, l'utilisation du tableur permet d'entraîner les élèves à exploiter un tel fichier. On peut, notamment, les amener à faire des choix (filtrer les données) construire des tableaux croisés de deux caractères, représenter les données par des graphiques adaptés... Dans cette ressource on s'intéresse aux données concernant l'académie de Grenoble, en 2021.

Chaque enseignant peut choisir de travailler à partir des données relatives à son académie ou tout autre territoire et extraire le fichier source Excel correspondant en ayant recours aux filtres proposés sur l'[open data du ministère de l'enseignement supérieur et de la recherche](#).

Le travail d'interprétation des résultats obtenus demeure primordial afin de mettre en exergue le rôle des statistiques et du traitement des données dans l'analyse de l'information chiffrée.

¹ Sitographie :

<https://www.data.gouv.fr/fr/>

<https://www.cnil.fr/professionnel>

<https://www.cnil.fr/fr/definition/open-data>

Scénario pédagogique

Modalités

Pour réaliser cette activité, il est préférable que les élèves disposent d'un ordinateur permettant ainsi de manipuler les données dans un tableur (salle informatique, classe mobile d'ordinateurs...). Il est également possible de proposer l'activité en salle « ordinaire », en vidéo-projetant l'écran du poste du professeur et en dialoguant collectivement sur les manipulations à effectuer.

L'activité peut être traitée individuellement ou en groupes.

Elle est conçue pour une séance d'une durée approximative de 1 h 30.

Déroulement

Les questions proposées illustrent la variété des manipulations possibles sur une quantité conséquente de données. Elles comprennent également des exemples d'analyse qui peuvent en découler.

Situation de l'activité

Le fichier source pour cette activité peut être téléchargé [ici](#).

On s'intéresse à la poursuite d'étude des néo bacheliers de la session 2021 dans l'académie de Grenoble.

Étape préalable : détermination d'un sous-ensemble d'individu, construction d'un tableau croisé.

La première étape consiste à construire le tableau croisé à deux caractères correspondant à la situation considérée.

Pour faciliter l'analyse, les données relatives aux poursuites d'études des néo bacheliers (fichier source : colonnes BC à BF) peuvent être reportées sur une deuxième feuille dans un tableau donnant les effectifs pour chaque type de formation que l'on souhaite étudier (fichier source : colonne K).

	Effectif des admis néo bacheliers	Admis issus de voie générale	Admis issus de voie technologique	Admis issus de voie professionnelle
Autre formation	763	440	223	100
BTS	4207	711	1787	1709
BUT	2949	1686	1159	104
CPGE	1327	1256	57	14
Ecole d'ingénieur	472	439	33	0
Ecole de commerce	113	87	23	3
EFTS	71	27	26	18
IFSI	674	228	240	206
Licence	6169	5308	624	237
Licence_LAS	408	356	50	2
PASS	1268	1247	20	1

Pour obtenir les effectifs on utilise l’instruction :

=SOMME.SI(Sheet!\$K:\$K;Feuil2!\$A2;Sheet!BC:BC)

Elle permet d’afficher la somme des effectifs de la colonne BC (effectif des admis néo bacheliers) du fichier source (onglet *Sheet*) en ne tenant compte que des étudiants de la catégorie « autre formation » (cellule A2 du nouvel onglet *Feuil2*), parmi les formations indiquées colonne K du fichier source.

Les autres cellules du tableau sont ensuite complétées grâce à l’outil de recopie du tableau.

Exemples de questionnements

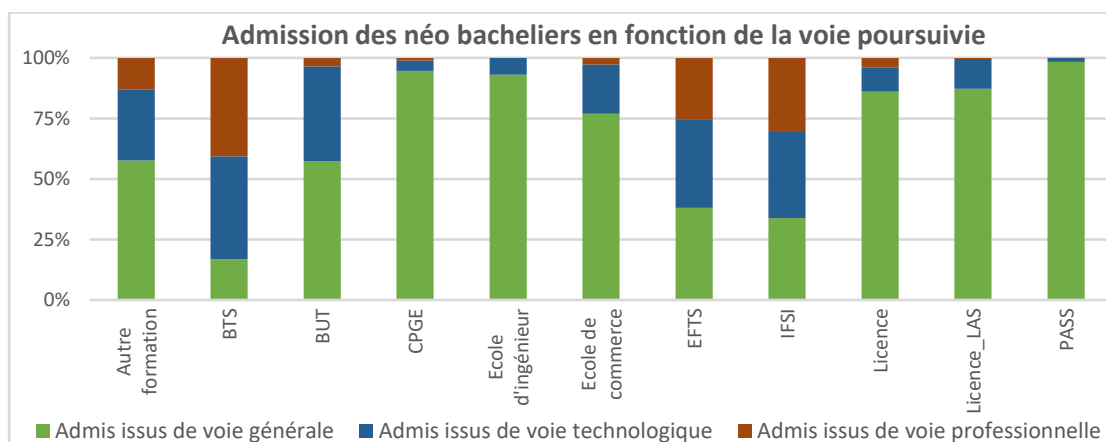
Dans un premier temps, on peut proposer des questions pour lesquelles l’analyse graphique des données est efficace.

1. Le choix post-bac de la poursuite d’étude est-il impacté par le type de baccalauréat obtenu (général, technologique ou professionnel) ?

Un premier point est la vigilance par rapport aux effectifs dans chacune des voies. Ici il semble plus opportun de travailler en fréquences plutôt qu’en effectifs.

La représentation graphique permettant de visualiser le type de baccalauréat obtenu par ces élèves en fonction de la formation choisie amène à plusieurs constats :

- pour le public étudié, la licence PASS, les écoles d’ingénieur et de commerce, les licences, les CPGE accueillent presque exclusivement des néo bacheliers généraux ;
- le recrutement en écoles d’infirmiers (IFSI) est équivalent quel que soit le type de baccalauréat obtenu ;
- plus de 40 % des élèves admis en BTS sont issus d’un baccalauréat professionnel.



Retrouvez éducol sur



Dans un deuxième temps, les constats réalisés graphiquement pourront être confortés par une analyse quantitative basée sur des calculs de fréquences.

Les outils du tableur pourront être utilisés pour automatiser le calcul des fréquences marginales et des fréquences conditionnelles. Un travail est donc à mener sur le choix de la condition et sur les différentes analyses que ce choix permet de proposer.

L'analyse du profil des admis en BUT conduit à mobiliser les notions de fréquence marginale et de fréquence conditionnelles.

2. Parmi les néo bacheliers, les filières BUT accueillent-elles majoritairement des étudiants titulaires d'un baccalauréat technologique ?
3. Est-il exact que parmi les néo bacheliers généraux, moins d'un sur six poursuit en BUT ?

Ce type de questions est l'occasion de sensibiliser les élèves sur les quotas d'admission dans certaines voies, ici par exemple la voie technologique.

Avant de réaliser des calculs de fréquences, les élèves doivent prendre le temps de bien définir la population concernée. En effet sur le graphique de la question 1, on constate que plus de la moitié des admis en BUT sont issus d'un baccalauréat général.

La question 3 mobilise une fréquence conditionnelle : $\frac{1686}{11\,785} \approx 0,143$ soit moins de $\frac{1}{6} \approx 0,166$.

Ainsi, si 57,2 % des néo bacheliers admis en BUT sont titulaires d'un baccalauréat général, moins d'un néo bachelier de la voie générale sur 6 s'oriente en BUT.

On peut construire avec les élèves les tableaux de fréquences conditionnelles suivants et envisager d'autres questions, en particulier en fonction du profil des élèves et du parcours d'études qu'ils envisagent.

	Part des lycéens issus de voie générale parmi les admis dans cette formation	Part des lycéens issus de voie technologique parmi les admis dans cette formation	Part des lycéens issus de voie professionnelle parmi les admis dans cette formation	Total
Autre formation	57,7	29,2	13,1	100
BTS	16,9	42,5	40,6	100
BUT	57,2	39,3	3,5	100
CPGE	94,6	4,3	1,1	100
École d'ingénieur	93,0	7,0	0,0	100
École de commerce	77,0	20,4	2,7	100
EFTS	38,0	36,6	25,4	100
IFSI	33,8	35,6	30,6	100
Licence	86,0	10,1	3,8	100
Licence LAS	87,3	12,3	0,5	100
PASS	98,3	1,6	0,1	100

	Part des lycéens issus de voie générale parmi les admis dans cette formation	Part des lycéens issus de voie technologique parmi les admis dans cette formation	Part des lycéens issus de voie professionnelle parmi les admis dans cette formation
Autre formation	3,7	5,3	4,2
BTS	6,0	42,1	71,4
BUT	14,3	27,3	4,3
CPGE	10,7	1,3	0,6
École d'ingénieur	3,7	0,8	0,0
École de commerce	0,7	0,5	0,1
EFTS	0,2	0,6	0,8
IFSI	1,9	5,7	8,6
Licence	45,0	14,7	9,9
Licence LAS	3,0	1,2	0,1
PASS	10,6	0,5	0,0
Total	100	100	100

Dans un troisième temps, on s'intéresse à l'impact des spécialités suivies en terminale sur l'orientation des élèves.

4. Pour un lycéen de l'académie de Grenoble qui ne choisit pas l'enseignement de spécialité mathématiques et qui souhaite intégrer une classe préparatoire, quelle est la doublette qui lui donne le plus de chance de concrétiser son vœu ?

Pour cela on travaille sur un nouveau fichier s'intitulant *Statistiques-formation-par-enseignement-de-specialite-2021* accessible sur le site [Bilan académique de Parcoursup 2021 | enseignementsup-recherche.gouv.fr](https://bilan.academique.deparcoursup.2021|enseignementsup-recherche.gouv.fr)

On peut tout d'abord comparer le nombre de demandes d'admission en classe préparatoires parmi les élèves ayant suivi ou pas l'enseignement de spécialité mathématiques en terminale.

CPGE					
Avec spé Maths			Sans spé Maths		
candidats inscrits	candidats avec proposition	propositions acceptées	candidats inscrits	candidats avec proposition	propositions acceptées
9180	4243	1089	2083	688	141

Ce tableau permet, en préambule de l'activité, de travailler sur des fréquences et des fréquences conditionnelles. Voici des études possibles :

- parmi les élèves ayant reçu une proposition d'admission en classe préparatoire, la fréquence des élèves ayant suivi l'enseignement de spécialité mathématiques est $\frac{4\,243}{4\,243 + 688} \approx 0,86$;
- parmi les élèves ayant demandé une classe préparatoire, la fréquence des élèves ayant reçu une proposition d'admission est de $\frac{4\,243 + 688}{9\,180 + 2\,083} \approx 0,44$;
- parmi les élèves qui avaient choisi la spécialité mathématiques et demandé une classe préparatoire, la fréquence des élèves ayant reçu une proposition d'admission est de $\frac{4\,243}{9\,180} \approx 0,46$;

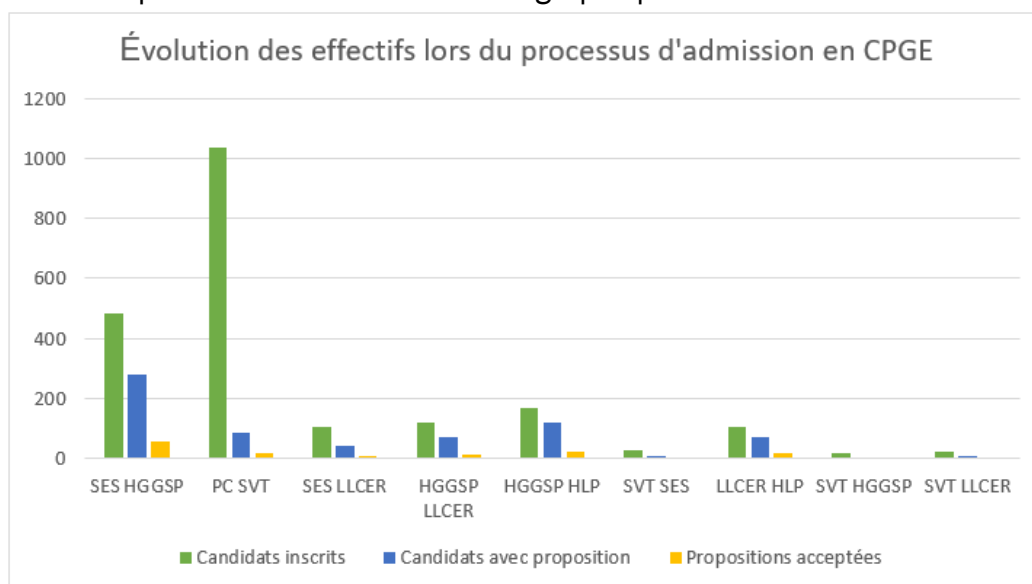
- parmi les élèves qui n'avaient pas suivi la spécialité mathématiques et avaient demandé une classe préparatoire, la fréquence des élèves ayant reçu une proposition d'admission est de $\frac{688}{2083} \approx 0,33$.

On peut ensuite regarder plus en détail les poursuites d'étude en classe préparatoire des élèves n'ayant pas suivi l'enseignement de spécialité mathématiques en terminale.

En sélectionnant les principales doublettes de spécialités ne comportant pas de mathématiques, on extrait un tableau dans lequel figurent les effectifs lors des trois phases du processus Parcoursup : vœux du candidat, proposition des formations et acceptation de la proposition par le candidat :

CPGE			
	Candidats inscrits	Candidats avec proposition	Propositions acceptées
SES HGGSP	485	277	57
PC SVT	1035	87	18
SES LLCER	107	41	6
HGGSP LLCER	118	73	14
HGGSP HLP	169	119	22
SVT SES	27	9	2
LLCER HLP	107	70	19
SVT HGGSP	15	5	1
SVT LLCER	20	7	2
TOTAL	2083	688	141

Ce tableau peut être illustré à l'aide du graphique suivant.



Cependant les effectifs étant très variables selon les choix de doublettes, un travail sur les fréquences paraît plus opportun pour une analyse plus fine.

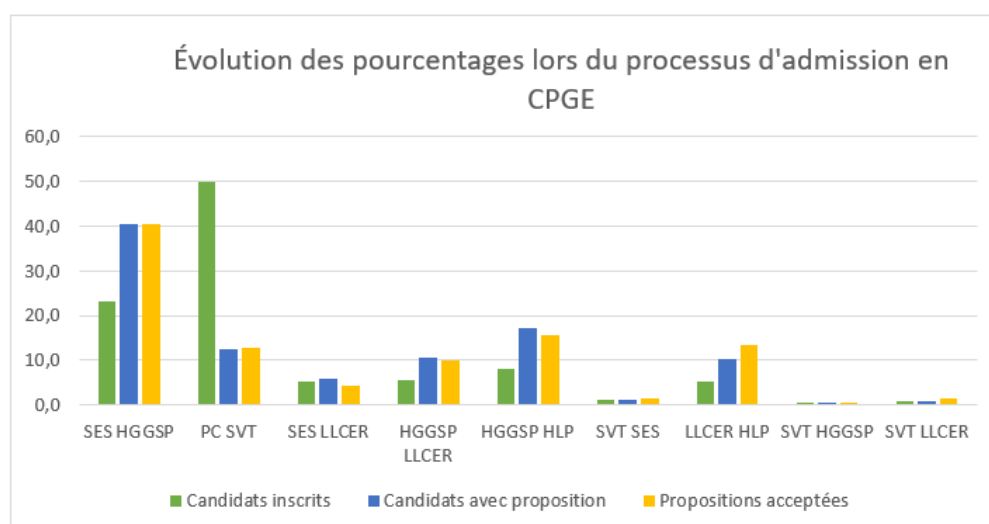
On dresse ainsi le tableau suivant :

CPGE			
	Candidats inscrits	Candidats avec proposition	Propositions acceptées
SES HGGSP	23,3	40,3	40,4
PC SVT	49,7	12,6	12,8
SES LLCER	5,1	6,0	4,3
HGGSP LLCER	5,7	10,6	9,9
HGGSP HLP	8,1	17,3	15,6
SVT SES	1,3	1,3	1,4
LLCER HLP	5,1	10,2	13,5
SVT HGGSP	0,7	0,7	0,7
SVT LLCER	1,0	1,0	1,4
TOTAL	100,0	100,0	100,0

On peut ainsi remarquer que quasiment 50 % des demandes en CPGE d'élèves n'ayant pas suivi l'enseignement de spécialité mathématiques en terminale proviennent d'élèves ayant suivi la doublette PC SVT.

Cependant, en proportion, les propositions sont davantage retenues par les formations pour les élèves de profil SES HGGSP, élèves qui représentent plus de 40 % de l'ensemble des propositions formulées par les formations. Les élèves de ce profil constituent également plus de 40 % des propositions acceptées.

Le graphique suivant illustre le tableau. Il convient de faire remarquer que, pour chaque doublette, les pourcentages ne se réfèrent pas aux mêmes populations.



Pour aller plus loin, on peut questionner la part de classes préparatoires comportant un enseignement de mathématiques dans l'académie de Grenoble et mettre en relation cette information avec l'étude réalisée.

Retrouvez éducol sur



Pour cela, on reprend le fichier Parcoursup, on applique un filtre à la colonne K afin de sélectionner seulement les « CPGE » puis on réalise le tableau du nombre de places disponibles en fonction de la filière.

Filières CPGE	Nombre de places
BCPST	96
ECG - Mathématiques appliquées + ESH	83
ECG - Mathématiques appliquées + HGG	86
ECG - Mathématiques approfondies + ESH	79
ECG - Mathématiques approfondies + HGG	82
ECT - Option technologique	54
Lettres	144
MP2I	48
MPSI	354
PCSI	375
PTSI	150
TSI	60

Enfin, on peut amener les élèves à s'interroger sur le besoin de données complémentaires qui pourraient permettre une étude plus complète comme par exemple le tableau croisé présentant les effectifs des élèves qui choisissent une filière CPGE proposant un enseignement de mathématiques en fonction des doublettes de spécialités suivies en terminale.

Commentaires de l'activité

Analyse a priori

L'activité d'analyse d'informations chiffrées s'appuie sur des manipulations effectuées ou projetées depuis un tableur. Elle offre l'occasion de présenter des instructions spécifiques notamment pour compter le nombre d'occurrences dans une plage de données. Néanmoins, aucune nouvelle compétence technique, spécifique à la manipulation du tableur, n'est exigée des élèves.

Le recours à des graphiques ou à des calculs de fréquences, en particulier de fréquences marginales ou conditionnelles par exemple, permet ensuite de mettre en évidence des relations ou des propriétés entre les données en vue d'une analyse. Celle-ci nécessite un temps de réflexion de la part des élèves.

Les questions suggérées mobilisent tour à tour une interprétation graphique, la manipulation de filtres simples ou avancés, le calcul de fréquences conditionnelles ou marginales.

Retrouvez eduscol sur



Manipuler

La manipulation se fait à partir d'un tableur. La spécificité d'un fichier de grande taille interdit de travailler « à la main » ou en balayant du regard les valeurs données. Confronter les élèves à la lecture d'un tel fichier permet de faire ressentir le besoin d'élaborer des stratégies pour automatiser l'extraction de données. Il importe de ménager régulièrement des temps de recherche individuelle ou en groupe pour assurer la phase d'appropriation.

Il s'agit ici de manipuler pour raisonner mais également de raisonner pour manipuler de façon pertinente : l'extraction de données, le choix d'un graphique, la mise en évidence de relations entre les données permettent de dégager des conclusions. Une phase d'essais-ajustements peut être nécessaire pour l'obtention d'une représentation adaptée des données au traitement du problème posé. À l'inverse, il paraît difficile de s'engager dans une manipulation sans phase préalable de raisonnement.

Verbaliser

Compte tenu des difficultés sous-jacentes, il importe de faire régulièrement verbaliser les élèves sur la nature des calculs automatisés à l'aide du tableur. La verbalisation joue également un rôle important à l'occasion de l'analyse statistique.

L'alternance de phases de recherche individuelle ou en groupe et de bilans collectifs s'impose. L'activité nécessite en effet de mettre en place pour chaque question un temps d'échanges portant sur la restitution de la recherche, la comparaison de méthodes, l'explicitation de contradictions, de points d'étonnement par rapport à des a priori, de conclusions, la formulation d'hypothèses...

Abstraire

L'abstraction intervient notamment lors de la mise en évidence des relations entre des données : identifier des relations d'union, d'intersection, de négation (travail sur la logique), distinguer les notions de fréquence marginale et fréquence conditionnelle ainsi que la nécessité de recourir à l'une ou à l'autre.

Pistes de différenciation possible

La part d'initiative laissée aux élèves durant la phase de manipulation peut être modulée en fonction de leurs profils. On peut suggérer ou non un type de représentation des données adapté au traitement d'une question que certains auront à réaliser tandis qu'elle sera donnée par l'enseignant à d'autres.

On peut également envisager un travail collaboratif au cours duquel chaque groupe doit produire des représentations ou des calculs différents qui seront des supports variés pour une synthèse élaborée collectivement.

Dans le cadre d'une différenciation, il est possible de varier le type de questionnement de manière à mobiliser un nombre plus ou moins important de données. À titre d'exemple, les élèves moins à l'aise pourraient travailler sur la répartition filles-garçons en licence qui met en jeu moins de paramètres.

Des prolongements peuvent aussi être envisagés pour les élèves plus rapides : analyse des choix de poursuite d'études des élèves boursiers, analyse des choix de poursuite d'études en fonction des départements...

Pistes d'évaluation

Ce type d'activité peut être le support de diverses évaluations :

- évaluation écrite : d'un compte rendu, par exemple ;
- évaluation de la compétence communiquer à l'oral : individuellement ou en groupes, on teste la capacité des élèves à expliquer de manière argumentée leurs réponses aux questions. L'activité peut aussi donner lieu à un rendu vidéo ;
- évaluation de la compétence représenter via la réalisation d'un graphique pertinent...

Pour aller plus loin

À partir du fichier Parcoursup, plusieurs questions peuvent s'envisager. À titre d'exemples : la part des élèves de la voie technologique ou professionnelle en CPGE, l'origine géographique des admis, la part des élèves issus de la voie professionnelle en université, la comparaison entre les filières d'études courtes ou longues...

Passer de l'échelle académique à l'échelle nationale permet des réflexions plus complètes sur les enjeux de l'admission en études supérieures en France.

Des données relatives à d'autres thèmes comme la pollution, la production ou la consommation énergétique, que l'on peut trouver sur le [site de l'INSEE](#), permettent d'envisager d'autres études.

Retrouvez eduscol sur



Complément

L'instruction **=SOMME.SI(Sheet!\$K:\$K;Feuil2!\$A2;Sheet!BC:BC)**

La fonction SOMME.SI détermine la somme des valeurs dans une plage en correspondance avec un critère donné.

Sa syntaxe générale est **SOMME.SI(plage;critère;[somme-plage])** :

- la *plage* dans laquelle le critère sera recherché sera ici la colonne K de l'onglet Sheet : on l'indique par l'expression **Sheet!\$K:\$K** dans laquelle \$ fixe le nom de la colonne lors de la recopie automatique ;
- le *critère* de recherche peut être indiqué dans la formule mais ici on signale que c'est le contenu de la colonne A (fixée par \$) sur la même ligne (ici on écrit en ligne 2, le numéro suivra lors de la recopie automatique) avec l'expression **Feuil2!\$A2** ;
- si *somme-plage* est omis, ce sont les valeurs de *plage* qui sont additionnées. Ici la colonne K ne contient pas de valeurs et on veut additionner les effectifs contenus dans la colonne BC de la feuille Sheet avec l'expression **Sheet!BC:BC**. Pas de \$ ici pour que la colonne suive lors de la recopie automatique.

Une valeur de la colonne BC de *Sheet* en ligne *n* n'est comptabilisée dans la somme que si le critère écrit dans la cellule A2 de *Feuil2* est présent dans la colonne K de *Sheet* à la ligne *n*.

La fonction SOMME.SI.ENS

La fonction SOMME.SI.ENS détermine la somme des arguments qui répondent *simultanément* à plusieurs critères.

Sa syntaxe générale est semblable à celle de SOMME.SI, mais on commence par la plage avec les valeurs à additionner :

SOMME.SI.ENS(somme-plage;plage1;critère1;[plage2];[critère2]...)

Par exemple dans l'onglet *Sheet* la colonne B donne le statut (Public, Privé sous contrat d'association...) de l'établissement ; l'instruction pour comptabiliser les élèves admis néo bacheliers (effectifs en colonne BC) de l'enseignement public (indiqué en colonne B) en filière de formation BTS (colonne K) est :
=SOMME.SI.ENS(Sheet!BC:BC;Sheet!B:B;"Public";Sheet!K:K;"BTS")

Il est aussi possible de n'indiquer qu'un seul critère pour cette fonction. On voit que **=SOMME.SI.ENS(Sheet!BC:BC;Sheet!\$K:\$K;Feuil2!\$A2)** serait équivalent à **=SOMME.SI(Sheet!\$K:\$K;Feuil2!\$A2;Sheet!BC:BC)**.